(12) UK Patent Application (19) GB (11) 2 299 914 (13) A

(43) Date of A Publication 16.10.1996

(21) Application No 9607539.5

(22) Date of Filing 11.04.1996

(30) Priority Data
(31) 9507454 (32) 11.04.1995 (33) GB

(51) INT CL[6]
H04Q 11/04 , H04M 3/24

(52) UK CL (Edition O)
H4K KTH KTK

(56) Documents Cited
EP 0624992 A2 EP 0613272 A2 US 5436886 A

(58) Field of Search
UK CL (Edition O) H4K KTH KTK
INT CL[6] H04M , H04Q
ONLINE:WPI

(71) Applicant(s)
General DataComm Advanced Research Centre Limited

(Incorporated in the United Kingdom)

3 Olympic Business Centre, Paycocke Road, BASILDON, Essex, SS14 3EX, United Kingdom

(72) Inventor(s)
Trevor Jones
Richard Barnett

(74) Agent and/or Address for Service
Loven & Co
Claxlete House, 62 Clasketgate, LINCOLN, LN2 1JZ, United Kingdom

(54) Data network switch with fault tolerance

(57) An ATM data network switch having a plurality of slot controllers 11a-11f, each slot controller having at least one external data link 12/13 and being separately connected to two separate switch fabrics 14a,14b, each switch fabric comprising means for switching a data cell transmitted from any one of the slot controllers to any of the other slot controllers, characterised in that both of the switch fabrics are active at the same time and each slot controller comprises means for determining the availability of the data paths to all the other slot controllers through both switch fabrics and for selecting for each cell to be switched a data path through one or other of the switch fabrics according to the availability determined.

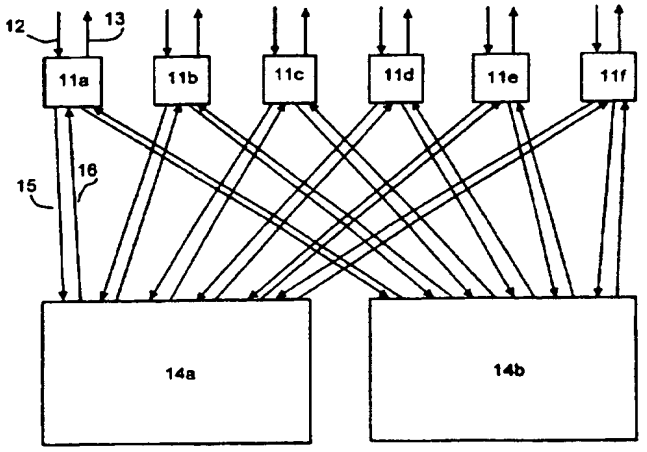

Fig 1

GB 2 299 914 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

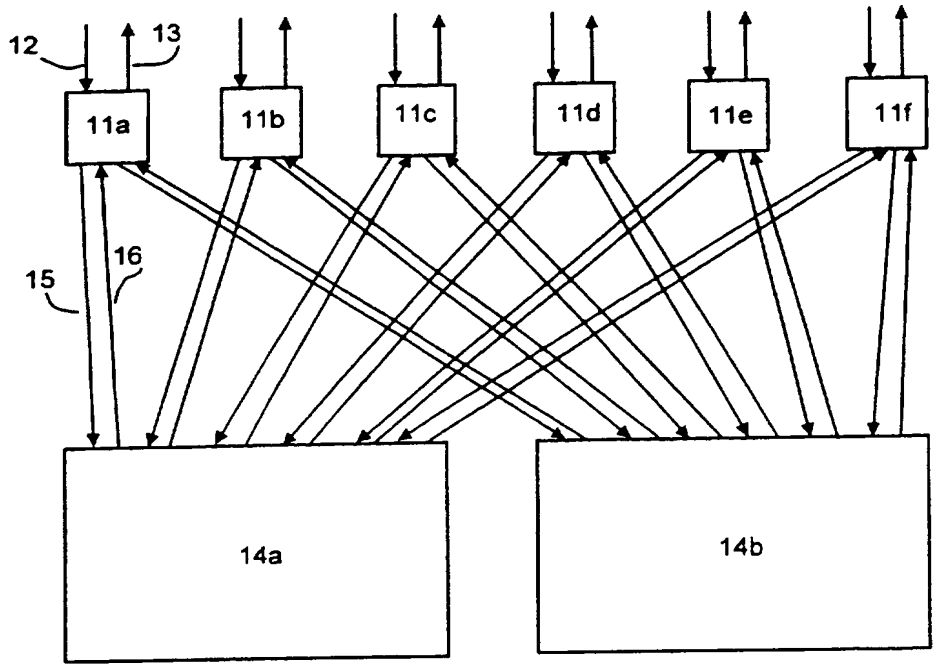

Fig 1

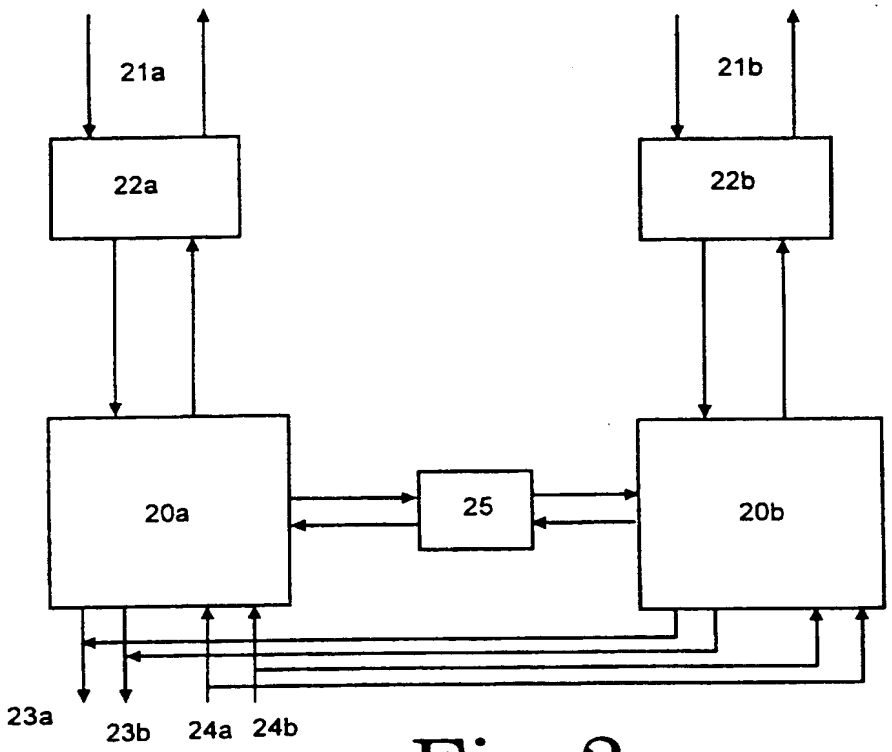

Fig 2

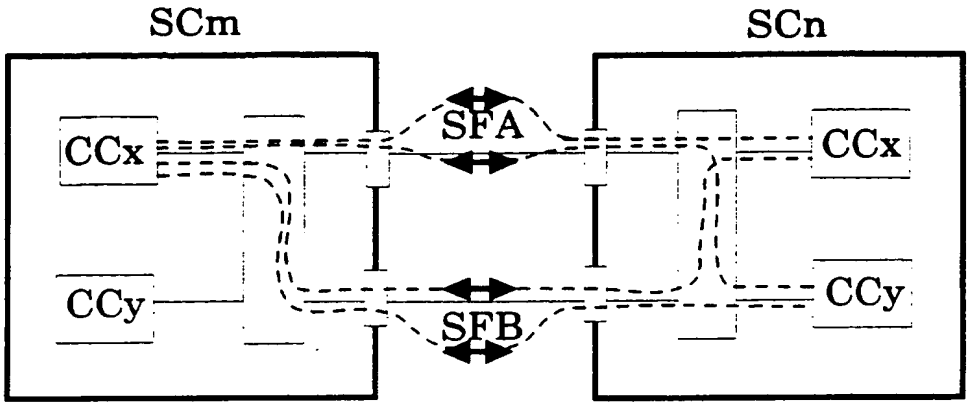

Fig 3

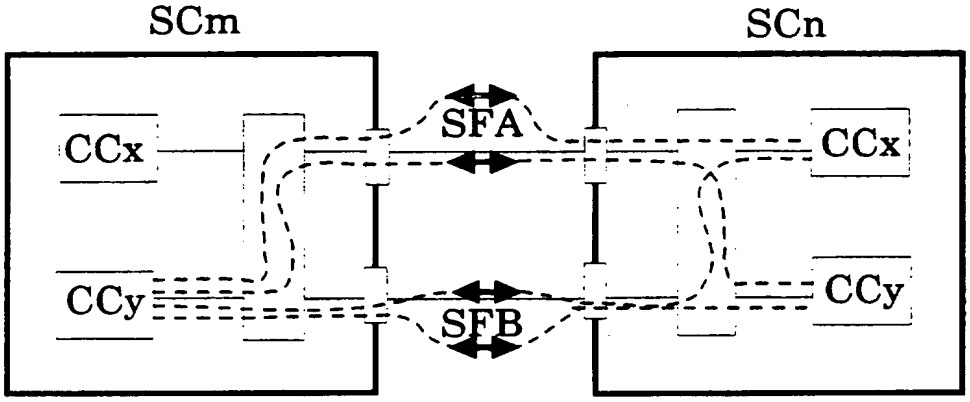

Fig 4

	Upper	Lower
0	Return Code	Reserved Slot / PRI / 0x3F / xy
1	Undefined	0xFF
2	Undefined	0xFF
3	n	n+1
4	n+2	n+3
5	n+4	n+5
26	n+46	n+47
27	n+48	n+49
28	n+50	n+51
29	n+52	checksum

Fig 5

	Upper	Lower
0	Return code	Returning Slot / PRI / 0x3E / xy
1	Undefined	Destination Slot Bitmask
2	Undefined	
3	n	n+1
4	n+2	n+3
5	n+4	n+5
26	n+46	n+47
27	n+48	n+49
28	n+50	n+51
29	n+52	checksum

Fig 6

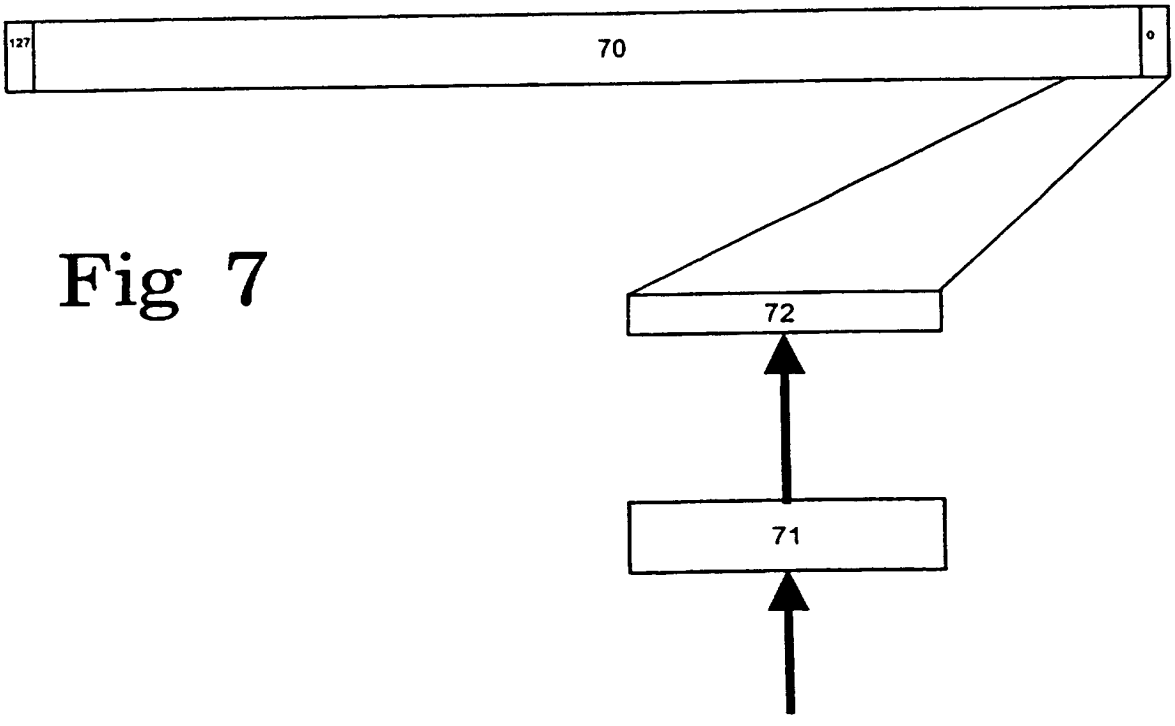

Fig 7

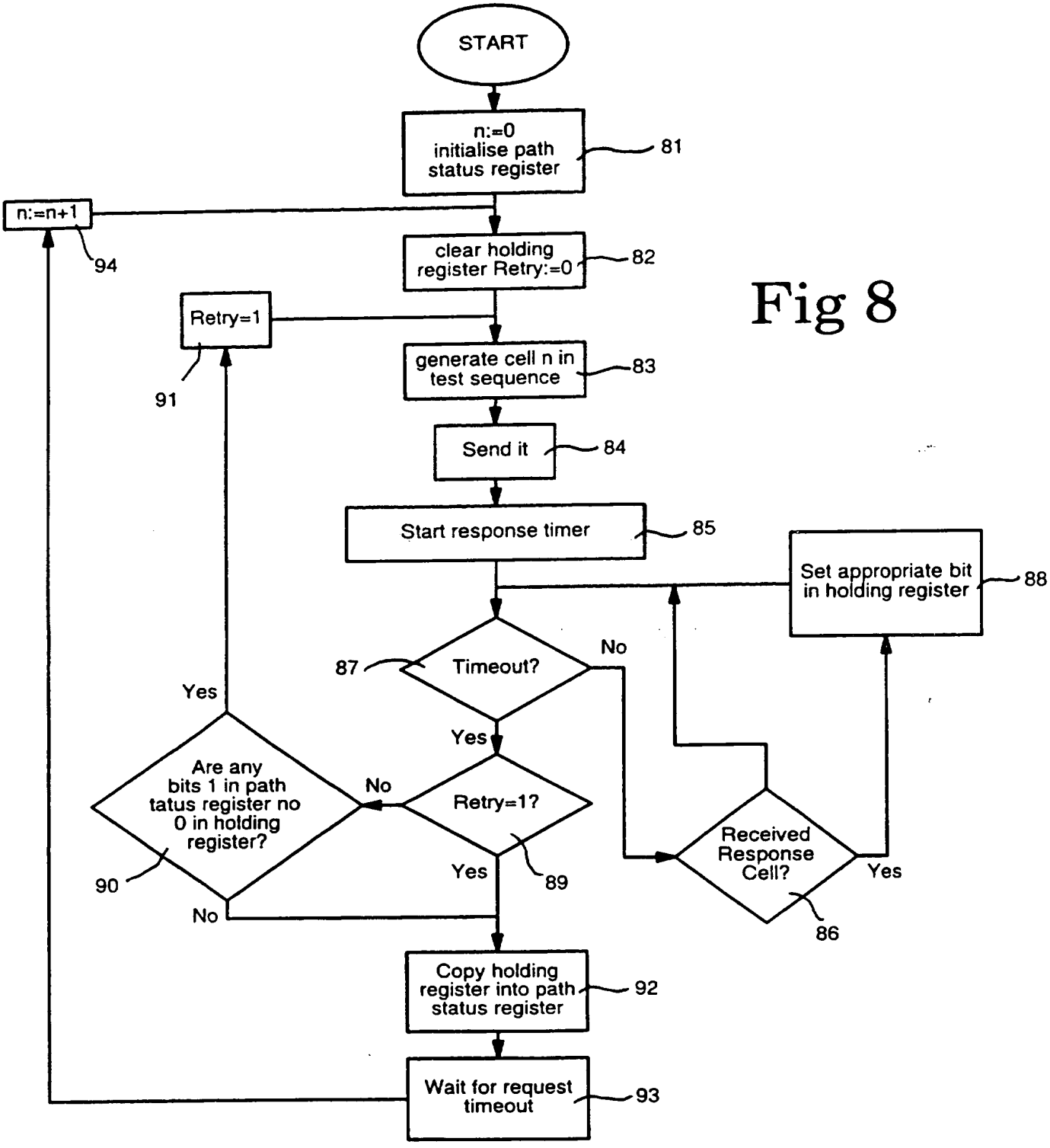

Fig 8

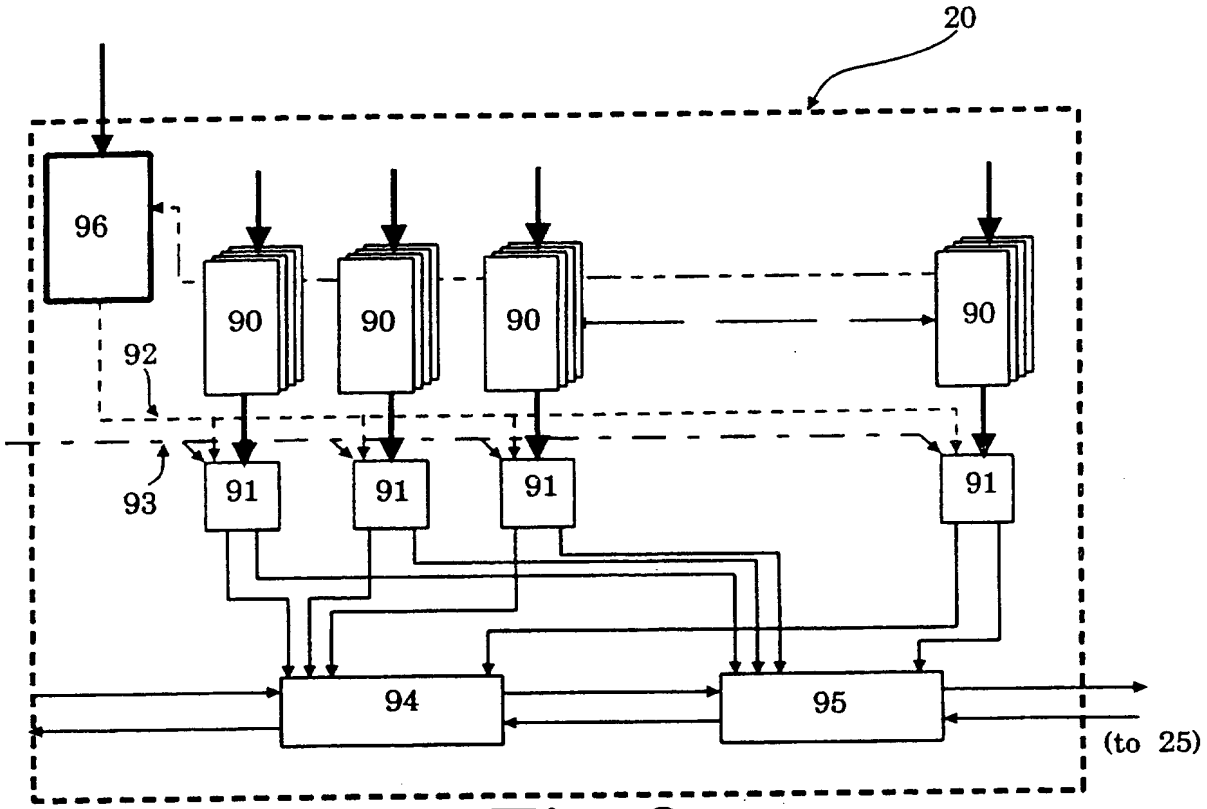

Fig 9

DATA NETWORK SWITCH WITH FAULT TOLERANCE

Field of the Invention

This invention relates to an Asynchronous Transfer Mode (ATM) data network switch for use in switching cells of data between a plurality of data links. The switch is arranged to have a high degree of tolerance to faults.

Background to the Invention

An ATM switch comprises, in general terms, a plurality of slot or link controllers each connected via an input port and an output port to a switch fabric, which is suitably a cross-point switch, to switch data cells from any input port to any output port. Each slot controller has a plurality of data links connected to it. The slot controllers comprise input controllers or receivers, whose principal function is simply to receive the bit stream from the external link and to divide it up into cells for presentation to the switch fabric, and output controllers or transmitters, which serve to convert the separate cells from the switch fabric into a continuous bit stream again for forwarding on the appropriate external link.

Since a fault in the switch fabric could cause failure of the complete switch, duplicate switch fabrics connected in parallel to the slot controllers are used. If a fault is detected in one switch fabric, switching is transferred to the second switch fabric, while the first is removed from use. It is possible to designate one of the slot controllers as a system controller arranged to monitor operation of the switch. For example, the system controller can send out "health check" cells to each other controller, to which the other slot controllers are arranged to respond by returning the cell to the system controller, which monitors the responses received. If the system controller does not receive all the responses, this may be due to a fault

in the switch fabric, and the system controller then switches from the first to the second switch fabric. This can result in cell loss.

A further problem with such an arrangement is that, although the switch fabric is fully duplicated, the second switch fabric remains inactive until it is required. It is therefore not possible to guarantee that the second switch fabric is fully operational when needed, since it can only be tested when in use. Further, no other advantage of duplication of switch fabrics is obtained. The capacity of the switch is identical with that of a switch having only a single switch fabric.

Summary of the Invention

According to the invention, there is provided an ATM data network switch having a plurality of slot controllers, each slot controller having a plurality of external data links thereto and being separately connected to two separate switch fabrics, each switch fabric comprising means for switching a data cell transmitted from any one of the slot controllers to any of the other slot controllers, characterised in that both of the switch fabrics are active at the same time and each slot controller comprises means for determining the availability of the data paths to all the other slot controllers through both switch fabrics and for selecting for each cell to be switched a data path through one or other of the switch fabrics according to the availability determined.

At least one of the slot controllers may comprise two or more cell processors, each connected to at least one external data link, and means for connecting each of the cell processors to each of the switch fabrics, to facilitate the handling of larger numbers of external links. The use of separate cell processors is a convenient way of increasing capacity in a slot controller, but it will be appreciated that by appropriate design of the cell

processor, greater numbers of external connections and internal switching paths may be provided for without the need for division into separate processors acting in parallel.

Preferably, each slot controller comprises means for periodically sending to each other slot controller via each switch fabric a "health check" data cell, means for receiving health check cells from other slot controllers and for returning each cell to its source via the same data path, and means for monitoring the return of health check cells from other slot controllers and for identifying therefrom the availability of individual data paths through each of the switch fabrics. The health check system establishes which paths are operating correctly.

It will be seen that, although the primary reason for providing two or more paths between each slot controller and each other slot controller is fault tolerance, it is also possible to use, say, two paths simultaneously to achieve, for example, 1.6 Gbps per slot rather than 800 Mbps for a single path, if full fault tolerance is not required. It also permits the assignment of different data transit priorities to the two paths, so that high priority data cells can pass through one path with minimal transit delay, while the bulk of the data cells, which are of lower priority and can tolerate greater transit delays, can pass through the other path.

Thus, in the dual redundant mode, there are two paths between each pair of slot controllers, through the two separate switch fabrics. A switch may, for example, support four classes of cell traffic, in descending order of priority: (1) CBR - Constant Bit Rate; (2) VBR - Variable Bit Rate; (3) ABR - Available Bit Rate; and (4) UBR - Unspecified Bit Rate; and each of these may have associated with it a switch fabric preference. For example, all traffic classes apart from CBR might be assigned a preference

for the first or A path, while CBR cells are given a preference for the second or B path. As long as the B path to the target slot controller is available for a particular CBR cell, it will use the B path, but if that path is unavailable, the control means in the source slot controller will automatically route the cell over the A path. Similarly, the other classes will reroute through the B path should the A path fail to a particular slot controller. The decision is made separately for each target slot controller from any particular source slot controller. Provided the total sustained rate is within the raw 800 Mbps capacity (for example) of a single switch fabric path, the slot controller will continue to operate at full load to any target slot controller provided at least one of the two paths is operating. Should both fail, the source slot controller is arranged to discard cells intended for the target slot controller.

One option in redundant mode would be to send, say, cells of priorities 1 and 3 through one switch fabric and those of priorities 2 and 4 through the other switch fabric, each switch fabric operating at a maximum of half of its maximum capacity, and therefore providing the possibility of re-routing cells through the other switch fabric should a path fail in the first, without the risk of exceeding the capacity of the switch to handle the total loading of all four priorities of cells.

In the double capacity mode, for example of 1.6 Gbps per slot, the full raw bandwidth of both switch fabric interfaces is available, although a single Virtual Connection (VC) is still limited to the throughput of one switch fabric interface in order to avoid re-sequencing cells. Should an inter-slot controller path fail, available throughput will halve between those two slot controllers as all of the cells have to be moved on to the same switch fabric interface. This is, of course, a considerable

improvement on the conventional arrangement where only one switch fabric is provided, or on the conventional redundant configuration, where the second switch fabric has to be switched in to replace the first, with resultant loss of throughput at the time of switch over. It will be appreciated that this mode of operation is really an issue of VC configuration rather than one of hardware, and therefore slot controllers within the same switch can operate in different modes according to demand.

Brief Description of the Drawings

In the drawings, which illustrate diagrammatically aspects of the construction and operation of an ATM cell switch according to exemplary embodiments of the invention:

Figure 1 shows the connections of the individual slot controllers to two switch fabrics in a simple switch;

Figure 2 shows an individual slot controller in more detail;

Figures 3 and 4 show possible data paths for a switch in which the slot controllers comprise a plurality of individual cell processors;

Figure 5 shows the structure of a health check request cell which can be transmitted through the switch fabric to determine data path availability;

Figure 6 shows the structure of a health check response cell returned by a slot controller in response to receipt of the request cell illustrated in Figure 5;

Figure 7 shows the logic within the slot controller handling the path status checking and recording;

Figure 8 is a flow diagram illustrating the operation of the health check algorithm; and

Figure 9 is a diagram illustrating the logic within the slot controller

controlling the selection of the output to one or other of the switch fabrics.

Detailed Description of the Illustrated Embodiment

Referring to Figure 1, the simple arrangement illustrated has six slot controllers 11a-f, each having external input and output links 12 and 13 respectively, and two separate switch fabrics 14a and 14b, each of a dynamic crosspoint type and having input and output connections 15 and 16 respectively to each of the slot controllers 11. The structure of the slot controllers is, for example, of the general type described and claimed in our earlier application GB9505358.3, and ATM cells arriving on an input link 12 may be processed in the general manner described in that application. Each slot controller comprises means for generating health check cells as hereinafter described, and for broadcasting the health check request cells to each other slot controller via both switch fabrics 14. In contrast with previous arrangements in which redundancy is provided, both switch fabrics remain active, rather than one being active and the other inactive until a failure in the first causes the second to be activated. Upon receipt of a health check request cell in a slot controller, a health check reply cell is generated and transmitted back to the source of the original request cell via the same data path. In this way, the originating slot controller receives reply cells from all the other slot controllers over the active data paths through the two switch fabrics, and can thereby determine the availability to itself of all the possible data paths in the switch. Each slot controller comprises memory in which the availability data can be stored so that each cell arriving at the slot controller from an external link can be routed within the switch according to the availability stored therein. For example, if in slot controller 11a the data path to slot controller 11d through switch fabric 14a is flagged as unavailable in the slot controller

memory, then a cell whose destination within the switch is controller 11d will be routed through the other switch fabric 14b.

Referring to Figure 2, each slot controller may optionally comprise two cell processors 20a and 20b, each in the form of an ASIC and having associated RAM defining input and output buffers, the processors also providing buffer management functions, to support two 622.08 Mbps links 21a and 21b, or up to 16 links at lower speeds, via physical interfaces 22a and 22b. The slot controller has two output connections 23 and 23b to the two switch fabrics A and B respectively, and two input connections 24a and 24b for cells returning from the two switch fabrics. An arbitration logic 25 controls the output from each cell processor 20 to the respective switch fabrics and input to the cell processors from the switch fabrics. When a cell processor wishes to send a cell to one or other of the switch fabrics, a request is sent by the cell processor to the arbitration logic. The mechanism by which the request is generated is described hereinafter with reference to Figure 9. The arbitration logic is arranged to simply to ensure that both cell processors are not sending cells to the same switch fabric at the same time. This is done by sending a grant signal back to the processor to permit it to send its cell. The processor cannot proceed until it has received the grant, and the grant is decided on the basis of alternation between the two cell processors when there is a conflict for the same switch fabric at the same time; in such an event, one of the cell processors has to wait to transmit its cell until the other has sent its cell.

Figures 3 and 4 illustrate the different paths between two separate slot controllers. With two cell processors in each slot controller and two switch fabrics, the number of paths which are available and which need to be checked is increased to eight, as follows:

- Slot controller SCm, cell processor CCx via SFA to slot controller SCn, cell processor CCx;
- Slot controller SCm, cell processor CCx via SFB to slot controller SCn, cell processor CCx;
- Slot controller SCm, cell processor CCx via SFA to slot controller SCn, cell processor CCy;
- Slot controller SCm, cell processor CCx via SFB to slot controller SCn, cell processor CCy;
- Slot controller SCm, cell processor CCy via SFA to slot controller SCn, cell processor CCx;
- Slot controller SCm, cell processor CCy via SFB to slot controller SCn, cell processor CCx;
- Slot controller SCm, cell processor CCy via SFA to slot controller SCn, cell processor CCy; and
- Slot controller SCm, cell processor CCy via SFB to slot controller SCn, cell processor CCy.

In addition, the switch fabrics may be arranged to handle cells of different priority in different ways, effectively creating a further diversification of paths. For example, in the crosspoint switch fabric used by the switch in accordance with the illustrated embodiments, the switching is carried out using ASICs which are configure to allow a cell to pass, or to block its passage, according to the switch fabric header in the cell. Part of the switching takes account of the different cell priorities which can be assigned to the cells, and cells of the different priorities are handled differently by the ASICs. Thus, if there is provision for two different priority classes through each switch fabric, there may effectively be sixteen different paths between each pair of slot controllers. Each of these paths has to

be checked for availability. In order to understand how this is done, it is necessary first to explain the operation of the health check system. The slot controllers continually check the paths to each other slot controller using health check request cells. These are special cells generated and checked by health check control means in the slot controllers to verify the availability of the data paths through the switch fabrics.

The structure of a health check request cell is illustrated in Figure 5. The low byte of the first word contains six bits of link code with the most significant bit being the priority bit and the least significant bit being the xy bit. The xy bit selects to which cell processor (CCx or CCy) the cell is to be routed. If it is set to 0, the cell goes to the CCx processor, and if it is 1, the cell goes to the CCy processor. The link code used for health check request cells is 0x3f ("0x" signifies a hexadecimal value). The upper byte of the first word is used to contain the source slot controller number (0x00-0x0f) in the lower nibble and the return codes in the upper nibble. Valid return codes are:

- 0x0. This means that the response cell should be returned using the SFA port and routed to the CCx cell processor;
- 0x1. This means that the response cell should be returned using the SFA port and routed to the CCy cell processor;
- 0x2. This means that the response cell should be returned using the SFB port and routed to the CCx cell processor; and
- 0x3. This means that the response cell should be returned using the SFB port and routed to the CCy cell processor.

A health check request cell has all bits of the Slot Controller Destination set to 1 (lower byte in word 1) to cause the cell to be broadcast to all slots. Our earlier published UK Patent Application No 2 273 224 discloses

a system of multicast distribution of ATM cells within an ATM Cell Switch, and this system is employed in the switch of the invention. The next 53 bytes consist of an incrementing sequence of bytes, based on a pseudorandom seed, to provide a payload for the cell. The actual values are not important to the functioning of the health check cell, the load merely serving to make the cell physically the same as normal payload cells. The last byte in the cell is an internal cell checksum to prove data integrity; an error in the checksum would indicate the possibility of a fault short of failure in the path over which the cell had travelled.

At the receiving slot controller, the health check control means generates a health check response cell in response to receipt of each health check request cell, and sends this back to the originating slot controller, and cell processor within it, over the same data path as the request cell to which it is responding. The structure of the response cell is illustrated in Figure 6. The lower byte of the first word contains the special health check response cell link code (0x3e in hexadecimal) with the priority bit in the most significant bit and the xy bit in the least significant bit. The upper byte of the first word contains the slot number of the slot controller sending the response cell in the lower nibble and the return codes (copied from the request cell) in the upper nibble.

The lower bytes of the second and third words contain the destination slot bit mask. The appropriate bit within this word is set so that the cell is routed to the sending slot of the request cell that caused the generation of the response cell (the sender's slot number was obtained from the upper byte of the first word of the health check request). The remainder of the cell is a separate incrementing sequence of bytes, the internal checksum being recalculated to reflect the new header contents.

The path status checking and recording logic is illustrated by Figure 7. A 128-bit path status register 70 stores the availability of each path in the switch, in terms of "good" or "bad", represented by 1 or 0. Each slot controller sends a health check cell not only to each of the other 15 slot controllers, but also to itself. Thus, the 128 bits are made up of 16 slot controllers x 2 cell processors per slot controller x 2 levels of priority x 2 switch fabrics. (The two levels of priority referred to are those by which the switch fabric itself operates. The ASIC elements within the switch fabric which perform the switching operation are programmed for convenience to operate with two priority levels. This is an arbitrary arrangement which is not essential to the operation of the invention.) A decode logic 71 receives the response cells and generates an address in a holding register 72 and generates the response bit to be stored therein. The holding register is a 16-bit register which stores the results of one set of tests for the 16 slot controllers and then transfers these results to the appropriate 16-bit region of the path status register 70, in readiness for the next set of tests. As explained in more detail hereinafter with reference to Figure 8, before the contents of the holding register are transferred to the appropriate region of the path status register, they are compared with the existing contents to determine whether any paths previously available are now indicated as unavailable. If a change in this way is detected (the opposite changes are not considered - a path is treated as available until the tests indicate otherwise), the set of tests is repeated once and the results transferred to the path status register, regardless of the results.

Figure 8 illustrates the algorithm by which the health check is carried out. The first step (81) is to clear the path status register to all 0s (all bad), or all 1s (all good), and the value of n is set to 0. In the next step

(82), the holding register is cleared, and the value of the Retry flag is set to 0. A priority 0 health check request cell as hereinbefore described is built at step (83), and this cell is sent (84) over the appropriate interface according to the destination and switch fabric codes included in it. The response timer is started (85), and if a valid response cell (i.e. one which has a valid checksum) is received (86) before the end of the timeout period (87), the appropriate bit is set by the decode logic 71 (Figure 7) in the holding register 72 (at 88). If the end of the timeout period is reached without receipt of a response cell, or if the response cell is received, and the retry flag is still 0 (89), a comparison between the content of the holding register 72 and the corresponding region of the path status register 70 is carried out (at 90), and if a change is detected, the retry flag is set to 1 (91), and the process is returned to step 83 to repeat the test. If there is no change, the holding register is copied (92) to the relevant region of the path status register 70, and the algorithm then waits (93) for the health check poll period to expire before incrementing n (94) and returning to step 82.

If at the end of the timeout test at step 87 the retry flag value is 1, the comparison between the contents of the holding register and those of the relevant region of the path status register is not carried out, and the process proceeds immediately to step 92.

The algorithm continues until the full set of path tests has been carried out, before starting again. The result is that each slot controller maintains a path status register that contains the availability of the paths to each of the other slot controllers.

The loss of a single health check request/response cell does not cause the path to go bad due to the retry process. Two in a row must fail

before a path is marked as down although only a single good cycle is enough to make the path available again.

Figure 9 shows the request mechanism within one of the two cell processors 20a and 20b in Figure 2 by which the requests to the arbitration mechanism 25 are generated. Only one such mechanism is illustrated for convenience, but each cell processor 20 will incorporate such a mechanism. The cell processor 20 comprises a plurality of sets of output FIFOs 90, one set for each of the other slot controller destinations in the switch, and each set consisting of four FIFOs, one for each of the cell priorities provided for by the switch. (It will be appreciated that fewer or more priorities can be accommodated by varying the number of FIFOs in each set 90.) The four priorities are treated as 2 higher and 2 lower priorities, thus giving simply two levels to be considered. Each FIFO within a set has a preference bit (for switch fabric A or switch fabric B) pre-set in RAM 96 in the cell processor 20 which can be changed according to the switch set-up. Each FIFO set 90 provides to the RAM 96 a 1-bit non-empty request signal if it contains any ATM cells to be sent and this is signalled to a respective logic element 91, along with the respective preference bit, on signal line 92. A signal on line 93 from the health check mechanism provides the status of the path through the two switch fabrics, indicating whether the path is good or not (i.e. available or not). The logic elements 91, shown separately for the sake of clarity of explanation in Figure 9, are in practice suitably carried out as logic functions by a microprocessor forming part of the control ASIC in the cell processor.

Each logic element 91 has two output request lines, one to an "A request" element 94 and one to a "B request" element 95. If a logic element 91 receives a request signal from its respective FIFO set 90 indicating

that a cell is waiting to be sent to the switch fabric, it generates a request according to the following:

If both SF paths are good, the request is for the preference;

If the preference path is good and the other path is bad, the request is for the preference;

If the preference path is bad, but the other path is good, the request is for the other path; and

If both paths are bad, requests are generated for both paths, resulting in cells being transmitted and, in consequence of the path failure, lost or discarded. This is necessary because cells must be emptied from the FIFO as soon as possible to avoid congestion upstream of the cell processor.

The A and B request elements 94 and 95 then determine which is the highest priority cell waiting to be sent at any instant and generate an external request to the arbitration logic 25, to be handled as hereinbefore described. When the arbitration logic 25 signals to the cell processor to send its cell, the request elements between them signal to the appropriate FIFO 90 to send its next cell to the switch fabric determined by the logic element 91.

CLAIMS

1. An ATM data network switch having a plurality of slot controllers, each slot controller having at least one external data link thereto and being separately connected to two separate switch fabrics, each switch fabric comprising means for switching a data cell transmitted from any one of the slot controllers to any of the other slot controllers, characterised in that both of the switch fabrics are active at the same time and each slot controller comprises means for determining the availability of the data paths to all the other slot controllers through both switch fabrics and for selecting for each cell to be switched a data path through one or other of the switch fabrics according to the availability determined.

2. A switch according to Claim 1, wherein at least one of the slot controllers comprises two or more cell processors each connected to at least one external data link, and means for connecting each of the cell processors to each of the switch fabrics.

3. A switch according to Claim 1 or 2, wherein each slot controller comprises:

means for periodically sending to each other slot controller via each switch fabric a "health check" data cell;

means for receiving health check cells from other slot controllers and for generating a response cell for each received cell to indicate receipt by the slot controller;

means for sending each response cell to the source of the received cell via the same data path; and

means for monitoring the receipt of health check response cells from other slot controllers and for identifying therefrom the availability of individual data paths through each of the switch fabrics.

4. A switch according to Claim 3, wherein each slot controller comprises means for storing an indicator of the availability of each of the data paths from and to the said slot controller.

5. A switch according to Claim 4, wherein the monitoring means comprises means for checking if a returned cell is not received over a previously-available path within a predetermined time after sending of the original health check cell, and for initiating transmission of a further health check cell over said data path, and means for changing the availability stored in the storage means for said path if a response to said further health check cell is not received within a further predetermined period.

6. A switch according to Claim 4 or 5, wherein the monitoring means comprises means for changing the stored indicator of the availability of an unavailable path if a response to a health check cell is received over that path.

7. A switch according to any of Claims 4 to 6, wherein the means for sending out the health check cells comprises means for generating a health check request cell comprising a code indicating the source of the cell, and means for broadcasting the same cell on all the data paths from the slot controller to all other slot controllers in the switch.

8. A switch according to any preceding claim, wherein each of said switch fabrics consists of a multiple crosspoint switch.

9. A switch according to any preceding claim, wherein, in the slot controllers, said means for selecting the data path comprises means for reading a cell priority bit in the header of each cell and means for selecting the data path according to the value of the cell priority bit, if more than one data path is available.

10. A switch according to any preceding claim, comprising more than two switch fabrics.

11. An ATM data network switch, substantially as described with reference to the drawings.

The Patent Office

18

Application No: GB 9607539.5 **Examiner:** Al Strayton
Claims searched: 1-11 **Date of search:** 4 July 1996

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): H4K: KTH; KTK

Int Cl (Ed.6): H04M, H04Q

Other: ONLINE: WPI

Documents considered to be relevant:

Category	Identity of document and relevant passage		Relevant to claims
X	EP 0 624 992 A2	(ATT) See abstract	1
X	EP 0 613 272 A2	(ATT) See whole document	1 AT LEAST
X,P	US 5 436 886	(MCGILL) See esp. col.3, l.57-col.4, l.39	1 AT LEAST

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.